Final Engagement Summary Report

Hourly Demand Response C&I Baseline Methodology Review

Engagement Initiated: April 2021

Background:

The baseline methodology used to assess the capacity and dispatch performance of Commercial & Industrial (C&I) Hourly Demand Response (HDR) resources has been an active topic of discussion between the IESO and the Demand Response (DR) stakeholder community since 2019. Specifically, DR stakeholders have taken issue with the impact the in-day adjustment factor (IDAF) portion of the baseline calculation has on measured performance, particularly for non-weather sensitive loads.

The purpose of the IDAF is to capture deviations in load consumption from the historic 'High 15-of-20' baseline on the day of an economic, emergency, or test activation to provide a more accurate measurement of what the HDR resource would have been consuming in absence of the activation[1]. The concept of a baseline as an approximation of a resource's consumption profile had a DR activation not taken place is a common approach to DR performance measurement in other North American energy markets, and the use of baseline adjustments to capture day-of realities in a customer load profile is considered essential to accurate performance assessment[2].

DR stakeholders have argued that particularly for non-weather sensitive loads, the IDAF limits the participant's operational flexibility on an activation day, as customers can be penalized, via a lowered baseline, for taking action prior to the start of the event – either because the resource takes multiple hours to ramp down, or the facility foregoes operations for the entire day in anticipation of needing to curtail later on (i.e. cancel a scheduled shift or production run). Stakeholders asserted the current baseline methodology with in-day adjustment is resulting in an inaccurate measurement of the capacity being provided by HDR resources, and recommended the IESO make the IDAF optional[3].

IESO held extensive internal discussion on allowing HDR resources to opt-out of the IDAF. These discussions concluded that allowing to opt-out was unacceptable as it would create a potential misalignment between desired resource behaviour based on the definition of the product IESO procures through the Capacity Auction and the measurement of resource performance. In other words, a load

[1] A detailed description of the C&I HDR baseline calculation can be found in Market Manual 5.5, Section 1.6.26.3.1.

[2] The Demand Response Baseline, ENERNOC Whitepaper (2009). See the Appendix of this document for a list of additional studies that informed the scope of the baseline review.

[3] See Demand Response Working Group November 12, 2019 presentation for details.

that had already curtailed consumption before an activation notice could be credited as fully delivering it's obligation despite providing no incremental curtailment in real-time for the hour the HDR resource was scheduled and dispatched for.

However, IESO agreed to establish an empirical understanding of the accuracy of the current baseline and launched the HDR baseline review engagement in April 2021. The review sought to assess the performance of the current "High 15 of 20 with in-day adjustment" baseline methodology for C&I HDR resources relative to a set of alternative methodologies, including the High 15 of 20 baseline with no in-day adjustment[4].

Baseline Review Objective:

The objective of the HDR baseline methodology review was to provide a statistical evaluation of how effectively different baseline methodologies predict load in absence of a DR event, in order to:

1. Respond to strong stakeholder interest for a review of the effectiveness of the C&I HDR baseline per the history of the issue described above;

2. Ensure assessment of DR resources participating in the IESO-Administered Markets (IAMs) is reflective of performance in alignment with the Capacity Auction design enhancement objective of Improving Performance; and

3. Inform discussions on the HDR participation model and capacity market enhancements (e.g. introduction of Qualified Capacity methodology, amendments to performance thresholds)

The analysis sought to answer the following research questions:

1. How well does the current baseline perform relative to a set of alternative baseline methods?
2. Is there rationale for introducing an opt-out of the in-day adjustment factor (IDAF)?
3. Is there rationale for introducing multiple C&I HDR baseline methods in Ontario?
4. Is there rationale for applying the baseline at the contributor, rather than the resource level?

Analysis Approach:

Historic HDR load data was leveraged to test the relative performance of 25 baseline and in-day adjustment methods, including the current High 15 of 20 with IDAF baseline (see Appendix for full list of baselines tested). Hourly HDR resource consumption data from January 1, 2018 to December 31, 2019 was used to predict load using each baseline on a set of days on which no DR curtailment occurred. On these days, actual load was known, meaning the estimated consumption could be compared to the known value in order to assess the predictive power of each baseline. Estimated load and actual load on each proxy event day was used to construct performance metrics[5], enabling a statistical comparison of the relative accuracy of each baseline compared to the status quo.

In total, 21 virtual HDR resources and two physical HDR resources were included in the analysis. Each baseline was estimated on roughly 475 non-event days, meaning each baseline was estimated and

[4] See Resource Adequacy April 22, 2021 presentation for details

[5] The three performance metrics were a Relative Root Mean Squared Error, Average Relative Error, and Relative Error Ratio.

compared to actual load over 10,000 times. Segmentation analysis was also undertaken to determine whether multiple baseline types should be made available to market participants based on load type.

Key Findings:

The key findings of the review are as follows:

- The status quo High 15-of-20 with In-Day Adjustment baseline method applied at the resource-level is an effective, accurate method for assessing C&I HDR resource performance relative to other options, estimating load with 4.6% accuracy on average
- Baseline performance is relatively consistent across seasonal, size, and load variability segmentation meaning there does not appear to be a strong justification for introducing multiple baselines at this time
- The IDAF overwhelmingly has a positive impact on baseline accuracy and integrity, in both resource- and contributor-level applications
- Removing the IDAF would, on average, result in an under-estimation of load, which would result in resources being measured as delivering less capacity than is actually curtailed
- Allowing an IDAF opt-out at the contributor-level may result in a minor accuracy improvement for some resources, but the impact of these accuracy gains is marginal in the context of established HDR performance dead-bands

The Resource Adequacy September 23, 2021 and December 15, 2021 provide further details on the analysis approach, key findings, and conclusions.

Stakeholder Response:

Accuracy of a contributor-level application of the baseline

Stakeholders asserted that a contributor-level application of the baseline is more accurate than a resource-level application because it limits the extent to which the in-day adjustment can lower the baseline on the day of an event. They state that with the resource-level application, a large contributor on outage can incorrectly discount the quantity of curtailment the other contributors in the portfolio deliver to the system during the activation.

Contributor data was collected from aggregators in order to test this assertion, resulting in 13 resources being assessed. The analysis assessed the relative accuracy of the current baseline to (a) contributor-level application of the current baseline with IDAF, (b) contributor-level application of the current baseline without IDAF, and (c) contributor-by-contributor application of the IDAF (i.e., applying the IDAF at the contributor-level only when it improves the contributor's baseline accuracy).

The contributor-level review found that moving to a contributor-level application increases accuracy only marginally in some cases. However, it can cause much larger decreases in baseline accuracy due to the 20% cap on the in-day adjustment.

For example, if a contributor with a 10 MW baseline is consuming 15 MW on the event day, the baseline should be adjusted upward by 5 MW. Due to the IDAF, the adjustment is capped at 2 MW

(120%), but if the contributor belongs to a 100 MW baseline resource, the entirety of the 5 MW upward adjustment is captured (5%) resulting in a more accurate baseline.

IESO committed to continue exploring the impact of large contributor outages on assessed performance through further stakeholder engagement.

The capacity value of HDR resources

After the final results of the review were presented, stakeholders suggested that the intent of the baseline should not represent what the resource would have been consuming in absence of the activation as this does not indicated the capacity value of the resource. Stakeholders advanced the position that any reduction in load consumption below historic consumption (e.g. unadjusted High 15-of-20 baseline) should be considered capacity delivery, regardless of whether in response to an activation or not. Stakeholders stated they do not think the analysis for the accuracy of the baseline calculation was aligned to the capacity product being procured from demand response and other resources through the Capacity Auction.

The IESO's position that a baseline should represent load in the absence of an activation is consistent with other North American system operators, and the North American Energy Standards Board's baseline definition ("a baseline is an estimate of the electricity that would have been consumed by a demand resource in the absence of a Demand Response Event").

The IESO defines capacity as a resource's maximum ability to provide energy or reduce load when required, and further defines the Capacity Auction capacity product as an energy market must-offer/bid requirement, obligating resources to make energy/curtailment available for real-time balancing during specified hours. The IESO's current baseline methodology and application of that methodology is aligned with these definitions.

Conclusion

In July 2022, the IESO launched a new Capacity Auction Enhancements engagement, to engage with stakeholders on enhancements to the IESO's Capacity Auction. During the August 25, 2022 engagement session and August 26, 2022 Technical Session, the IESO indicated its intent to conclude discussion on the Hourly Demand Response C&I Baseline Methodology Review and focus discussion on exploring the impact of large contributor outages on assessed performance.

Stakeholders interested in continuing to follow engagement on this topic should follow the Capacity Auction Enhancements stakeholder engagement.

Appendix:

Reference Studies

- PJM Empirical Analysis of Demand Response Baseline Methods, KEMA, 2011
- Proposed Changes to the Demand Response Baseline, ISO New England, 2015
- Estimating Demand Response Load Impacts: Evaluation of Baseline Load Models for Non-Residential Buildings in California, Lawrence Berkeley National Laboratory, 2008
- NYISO SCR Baseline Study, KEMA, 2014
- The Demand Response Baseline, Enernoc Whitepaper, 2009
- *Measurement and Verification for Demand Response,* Prepared for the National Forum on the National Action Plan on Demand Response: Measurement and Verification Working Group, 2013

Baseline Methodologies Assessed

Each baseline method listed in the left-hand column was tested with each of the in-day adjustment methods in the righ-hand column, for a total of 25 baselines tested.

Baseline Methods	In-Day Adjustment Methods
1. **High 15-of-20**: average load on the highest 15 of the last 20 suitable business days (SBD) 2. **High 4-of-5:** average load on the highest 4 of the last 5 SBD 3. **High 5-of-10:** average load on the highest 5 of the last 10 SBD 4. **Middle 8-of-10:** average load on the middle 8 of the last 10 SBD (i.e. Day 2 through 9 of the last 10 SBD) 5. **Mean 10-of-10:** average load on the last 10 SBD	1. **Unadjusted:** No IDAF (i.e. unadjusted) 2. **Current IDAF:** 3 hours, one hour prior to first dispatch hour; +/- 20% cap; scalar adjustment 3. **Shifted IDAF:** Current IDAF with shifted timeframe: 3 hours, 3 hours prior to first dispatch hour 4. **Uncapped, scalar IDAF:** Current IDAF (scalar adjustment) with no +/- 20% adjustment cap 5. **Uncapped, additive IDAF:** Current IDAF with no +/- 20% adjustment cap and additive adjustment rather than scalar adjustment

Additive Adjustment: A fixed kW adjustment to the 'X of Y' baseline, applied across all event time intervals
Scalar Adjustment: A % multiplier adjustment to the 'X of Y' baseline, applied across all event time intervals
Uncapped: No limit to the magnitude of an adjustment to the 'X of Y' baseline